Supplementary Information for

Structure and transport mechanism of human riboflavin transporters

Ke Wang[1,2*], Huiwen Chen[1,2*], Lili Cheng[3*], Jun Zhao[4*], Bo Huang[5*], Di Wu[2,6], Xin He[3], Yumeng Zhou[3], Yaxuan Yuan[3], Feng Zhou[5], Juquan Jiang[1**], Ligong Chen[3**], Daohua Jiang[2,6**]

[1] Department of Microbiology and Biotechnology, College of Life Sciences, Northeast Agricultural University, No. 600 Changjiang Road, Xiangfang District, Harbin 150030, China

[2] Beijing National Laboratory for Condensed Matter Physics and Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China

[3] School of Pharmaceutical Sciences, Beijing Frontier Research Center for Biological Structure, Tsinghua University, Beijing 100084, China

[4] Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences at Weifang, Weifang, Shandong 261000, China

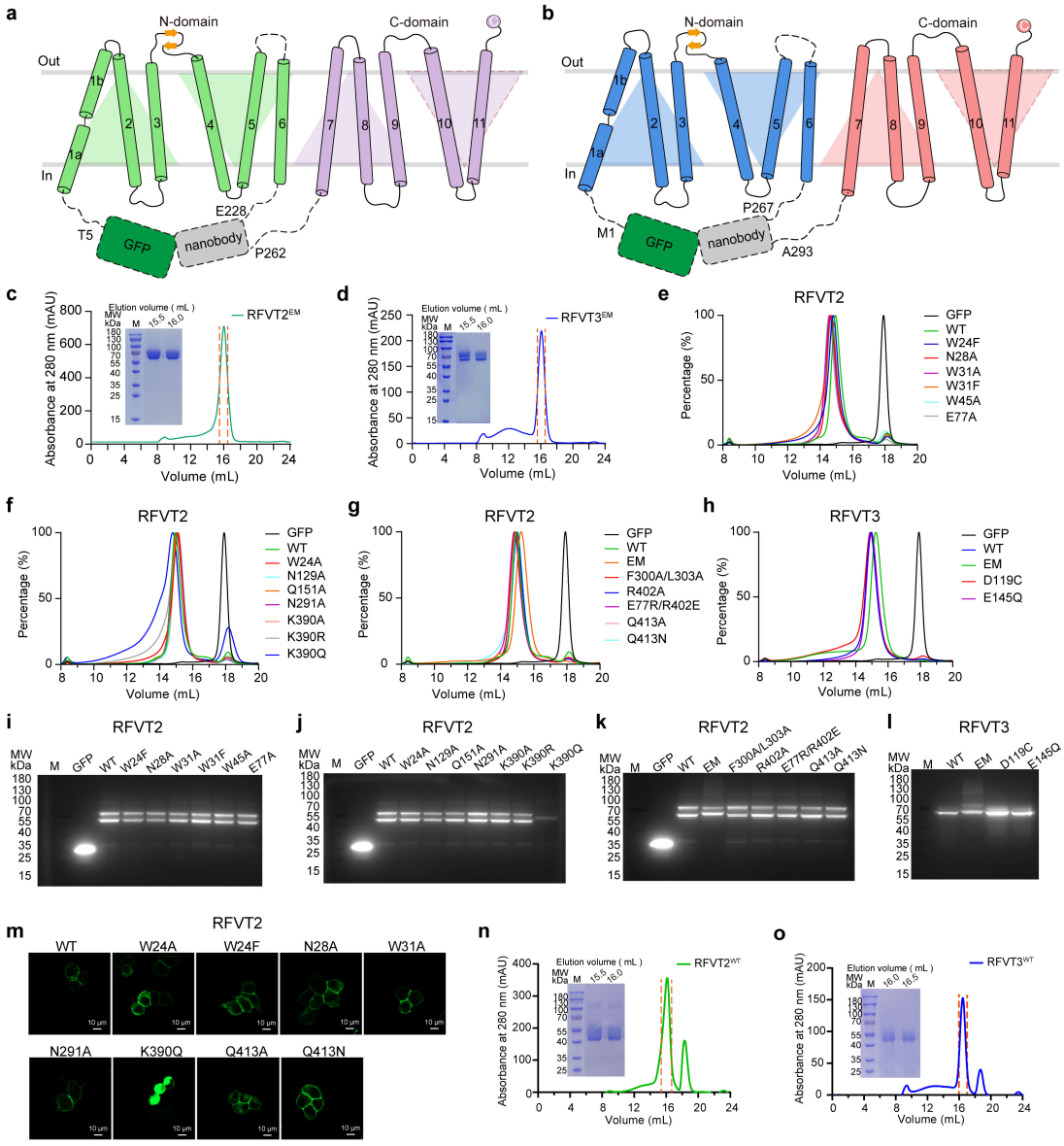
[5] Beijing StoneWise Technology Co Ltd., 15 Haidian Street, Haidian District, Beijing, China

[6] School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100190, China

* These authors contributed equally to this project.

** Correspondence emails: jiangdh@iphy.ac.cn (D.J.), ligongchen@tsinghua.edu.cn (L.Chen), and jjqdainty@163.com (J.J.)

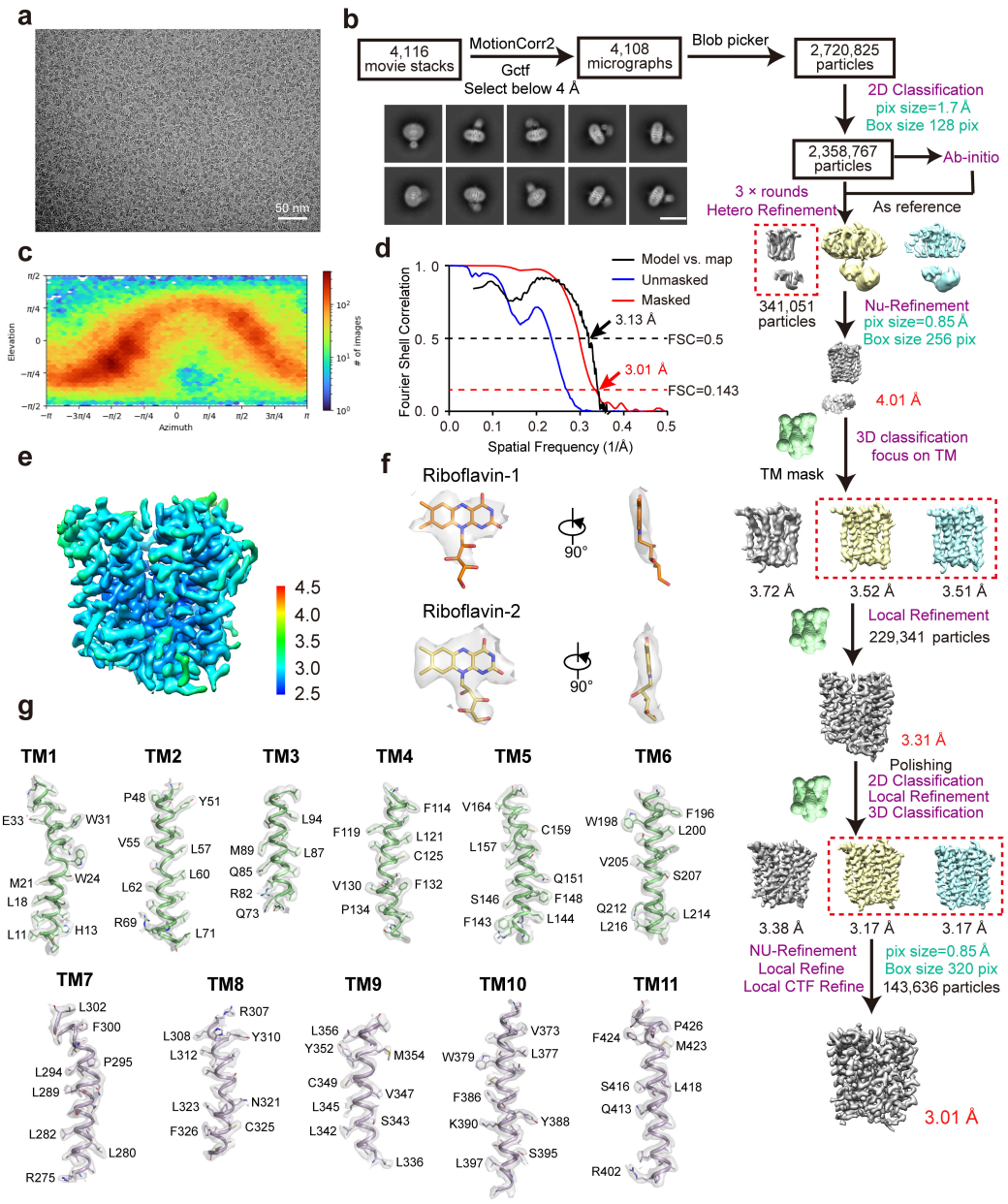
This document contains Supplementary Figures 1-9 and Supplementary Tables 1-2.



Supplementary Fig. 1. Purification of human RFVT2 and RFVT3.

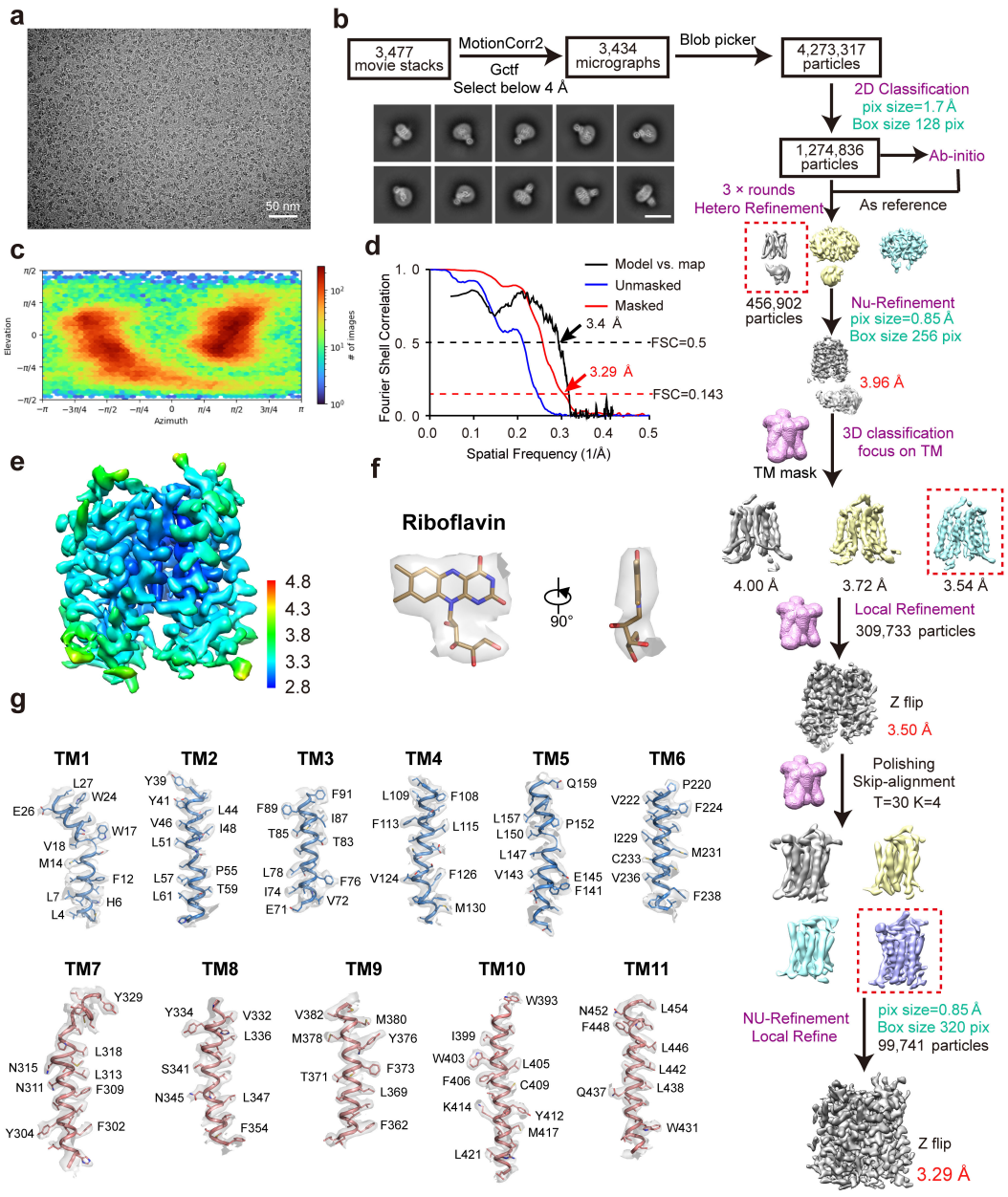
a,b, Topology of human RFVT2 (**a**) and RFVT3 (**b**). The N- and C-domains of RFVT2 are colored in green and purple; the N- and C-domains of RFVT2 are colored in blue and red. Green and gray dashed boxes represent GFP and nanobody against GFP respectively. Black dashed lines indicate the flexible regions which are not visible in cryo-EM maps. **c,d,** A representative size-exclusion chromatography (SEC) profile of purified human RFVT2 (**c**) and RFVT3 (**d**). Peak fractions used for cryo-EM analysis are indicated by red dashed lines. Peak fractions from SEC were visualized by SDS-PAGE with Coomassie blue staining. Red labels indicate samples used for cryo-EM analysis. The SEC profile and gel image are representative of 3 experimental replicates. **e-h,** Fluorescence-detection SEC (fSEC) analysis of WT RFVTs and their mutants. GFP serves as a control in black line. A Superose 6 Increase 10/300 GL column was used for fSEC analysis. **i-l,** In-gel fluorescence analysis of WT RFVTs and their mutants. SDS-PAGE gels were imaged using a Tanon 5200 Multi imager excited at 488 nm. **m.** Confocal images of WT RFVT2 and mutants expressed in HEK293 cells. Bar, 10 μm. **n,o,** A representative SEC

profile of purified human WT RFVT2 (**n**) and WT RFVT3 (**o**). Peak fractions used for proteoliposome reconstitution. Peak fractions from SEC were visualized by SDS-PAGE with Coomassie blue staining. The SEC profile and gel image are representative of 3 experimental replicates. Source data are provided as a Source Data file.



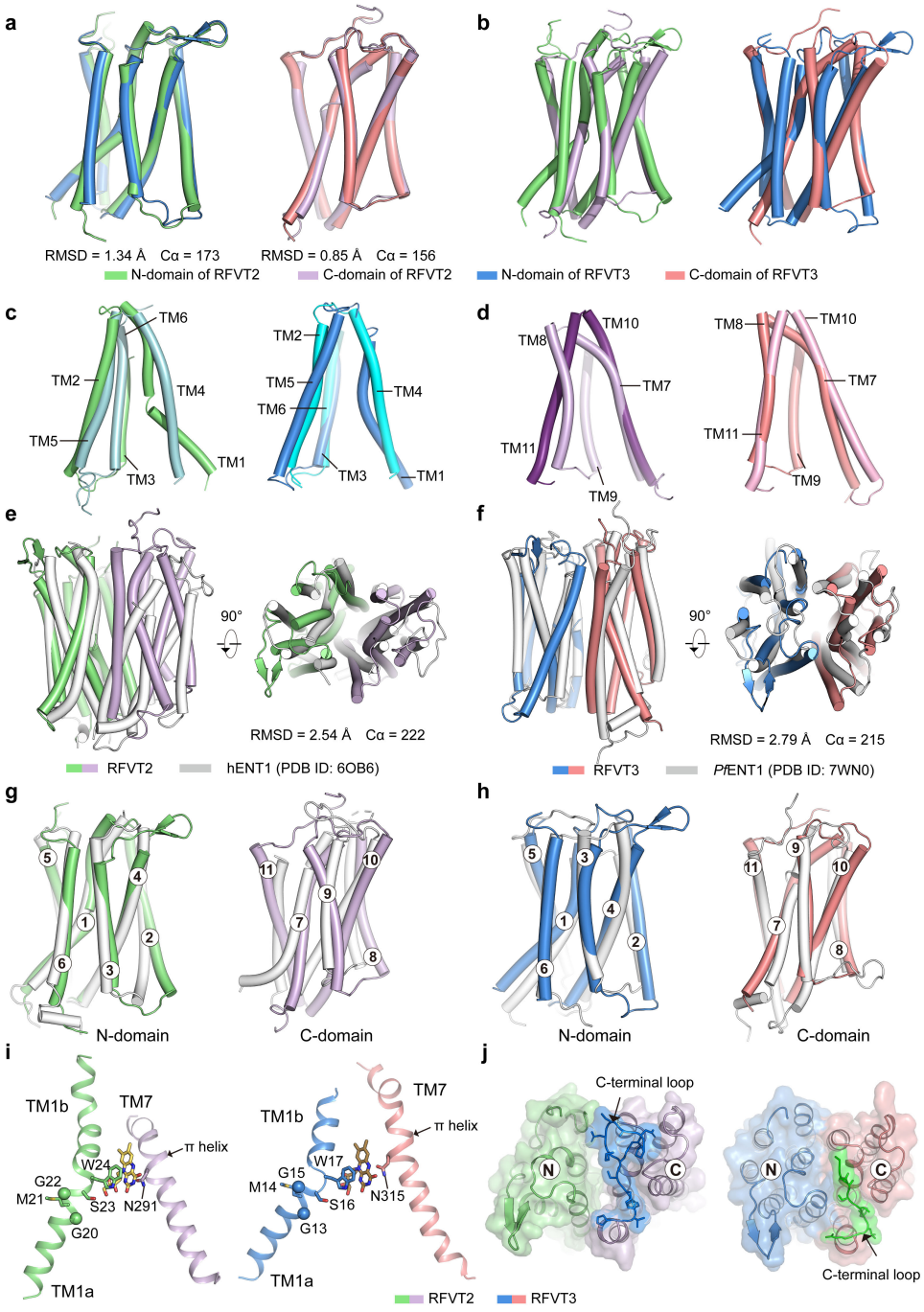
Supplementary Fig. 2. Cryo-EM data processing of human RFVT2.

a, A representative motion-corrected electron micrograph (out of 4,116 micrographs) of RFVT2. Scale bar, 50 nm. **b,** The flowchart for cryo-EM data processing of RFVT2. Scale bar in reference-free 2D class averages is 10 nm. **c,** Particle angular distribution for the final reconstruction. **d,** Fourier-shell correlation (FSC) analysis of RFVT2. The FSC curves of the final EM map without and with a solvent mask are colored in blue and red, respectively. Red arrows indicate the map resolution (FSC = 0.143). The FSC curve of the refined model fitting in the final EM map is colored in black. Black arrows indicate the model-map fitting resolution (FSC = 0.5). **e,** Local resolution map of RFVT2. **f,g,** EM densities for bound riboflavin (**f**) and TM helices (**g**).



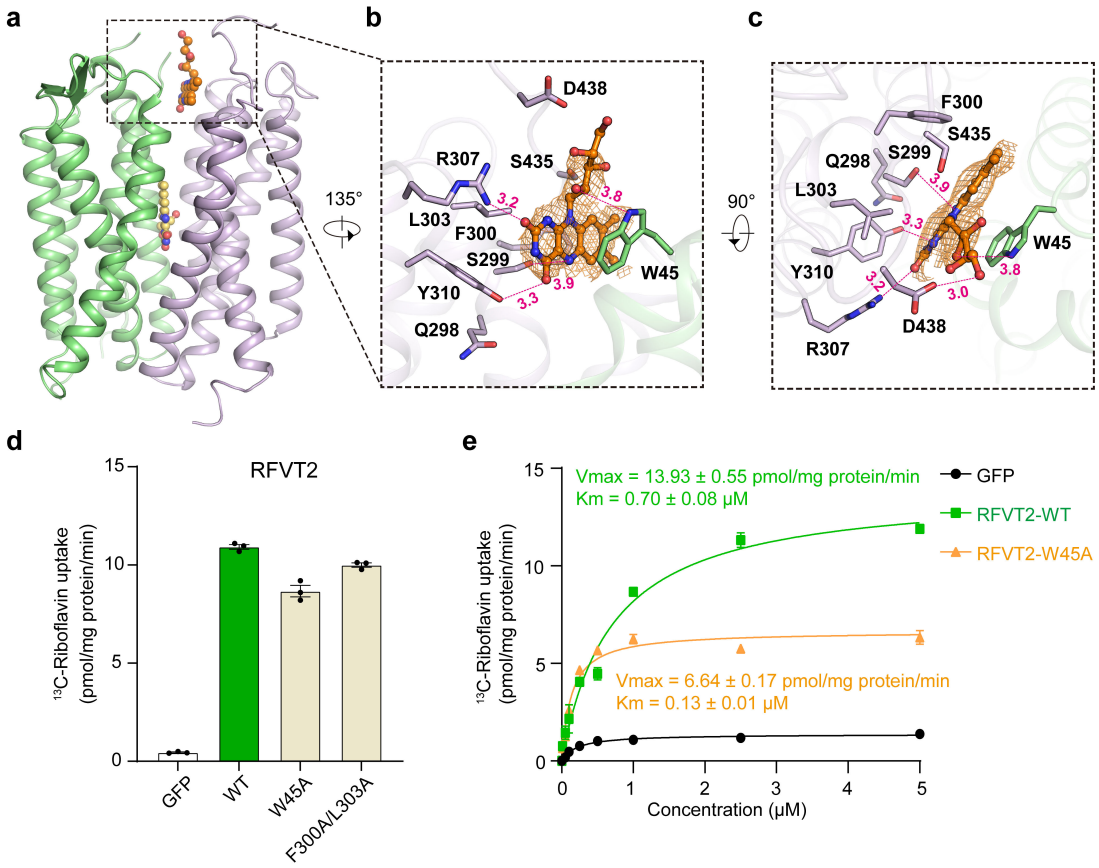
Supplementary Fig 3. Cryo-EM data processing of human RFVT3.

a, A representative motion-corrected electron micrograph (out of 3,477 micrographs) of RFVT3. Scale bar, 50 nm. **b,** The flowchart for cryo-EM data processing of RFVT3. Scale bar in reference-free 2D class averages is 10 nm. **c,** Particle angular distribution for the final reconstruction. **d,** Fourier-shell correlation (FSC) analysis of RFVT3. The FSC curves of the final EM map without and with a solvent mask are colored in blue and red, respectively. Red arrows indicate the map resolution (FSC = 0.143). The FSC curve of the refined model fitting in the final EM map is colored in black. Black arrows indicate the model-map fitting resolution (FSC = 0.5). **e,** Local resolution map of RFVT3. **f,g,** EM densities for bound riboflavin (**f**) and TM helices (**g**).



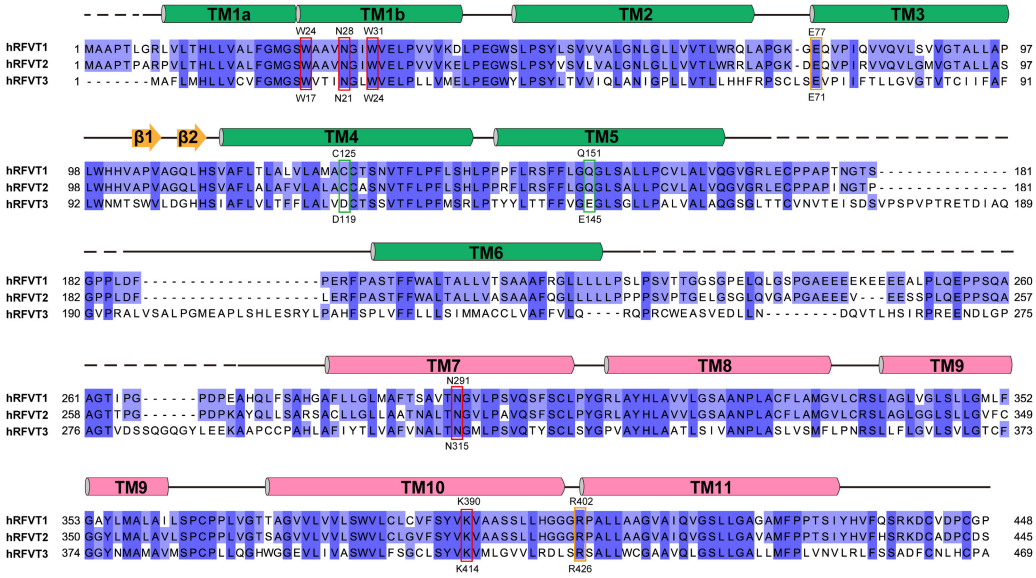
Supplementary Fig. 4. Structural features of RFVTs.

a, The superposition of the N-domain (left) and the C-domain (right) of RFVT2 and RFVT3. **b,** The superposition of the N-domain with the C-domain of RFVT2 (left) and RFVT3 (right). **c,d,** The superposition of TM1-3 with TM4-6 (**c**) and TM7-9 with TM10-11 (**d**) of RFVT2 (left) and RFVT3 (right). **e,f,** The superpositions of RFVT2 (**e**) with the outward-facing human ENT1 (PDB code: 6OB6) and RFVT3 (**f**) with the inward-facing *Plasmodium falciparum* ENT1 (PDB code: 7WN0). **g,h,** The superpositions of the N- and C-domain of RFVT2 (**g**) with those of human ENT1 and the N- and C-domain of RFVT3 (**h**) with those of *Plasmodium falciparum* ENT1. ENT1 in gray. **i,** The gating helices TM1 and TM7 in RFVT2 (left) and RFVT3 (right). **j,** The position of the C-terminal loop in RFVT2 (left) and RFVT3 (right).



Supplementary Fig. 5. The possible second riboflavin binding site in RFVT2.

a, Two riboflavin molecules bind to RFVT2. The N- and C-domains of RFVT2 are colored in green and purple. The upper and central riboflavin molecules are depicted in brown and yellow sticks. **b,c,** Detailed interactions between the upper riboflavin and RFVT2 viewed parallel to the membrane plane (**b**) and perpendicular to membrane plane (**c**). Red dashed lines indicate potential hydrogen bond interactions. **d,** [^{13}C]riboflavin uptake by WT RFVT2 and mutants. Cells expressing GFP serve as a control. Data are presented as mean ± s.e.m.; n = 3 independent replicates. **e,** Concentration-dependent [^{13}C]riboflavin uptake by RFVT2-W45A. Cells expressing GFP serve as a control. Data are presented as mean ± s.e.m.; n = 3 independent replicates. Source data are provided as a Source Data file.

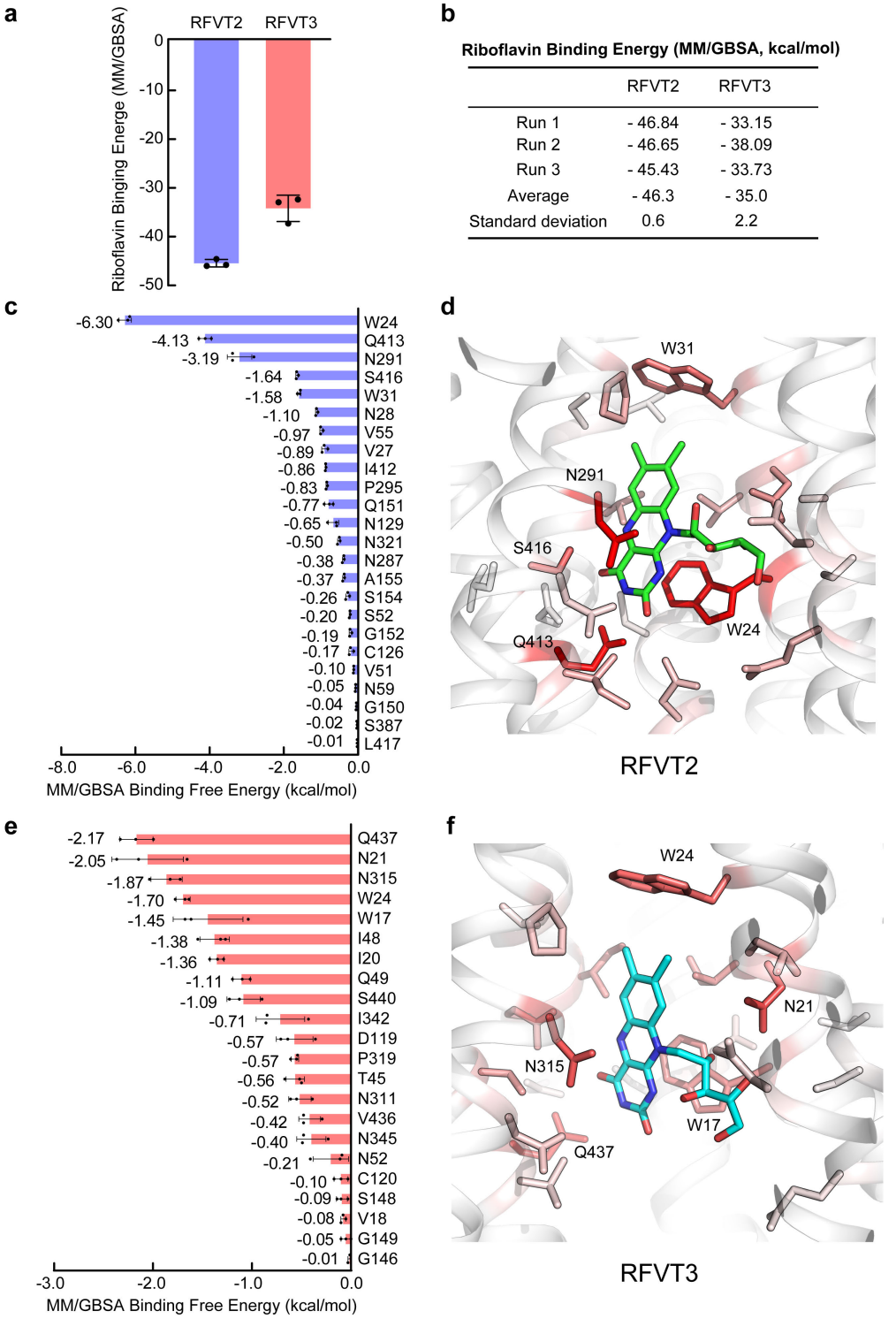


Supplementary Fig. 6. Sequence alignment of human RFVT1-3.

Sequence alignment of human RFVT1 (UniProt accession code: Q9NWF4), RFVT2 (UniProt accession code: Q9HAB3) and RFVT3 (UniProt accession code: Q9NQ40). Fully and partially identical residues are highlighted in dark blue and light blue. Secondary structure elements are labeled on the top of respective sequences. Red squares highlight residues involved in riboflavin binding. Orange squares highlight the salt-bridge interaction on the cytosolic side. Green square indicates the non-conserved E145 which determines proton coupling in RFVT3.

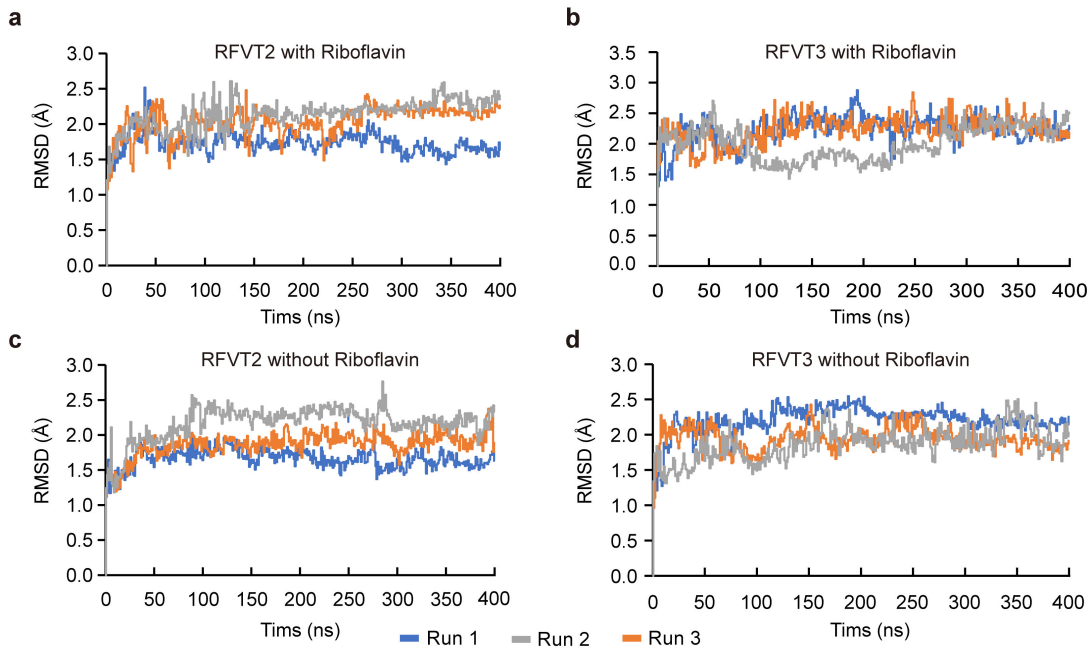
Riboflavin Binding Energy (MM/GBSA, kcal/mol)

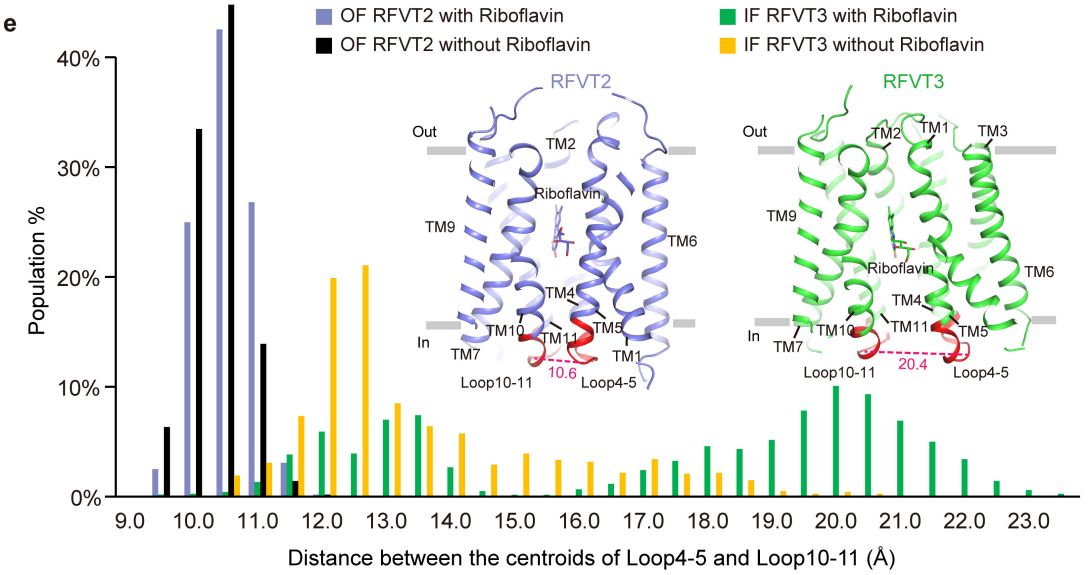
	RFVT2	RFVT3
Run 1	- 46.84	- 33.15
Run 2	- 46.65	- 38.09
Run 3	- 45.43	- 33.73
Average	- 46.3	- 35.0
Standard deviation	0.6	2.2

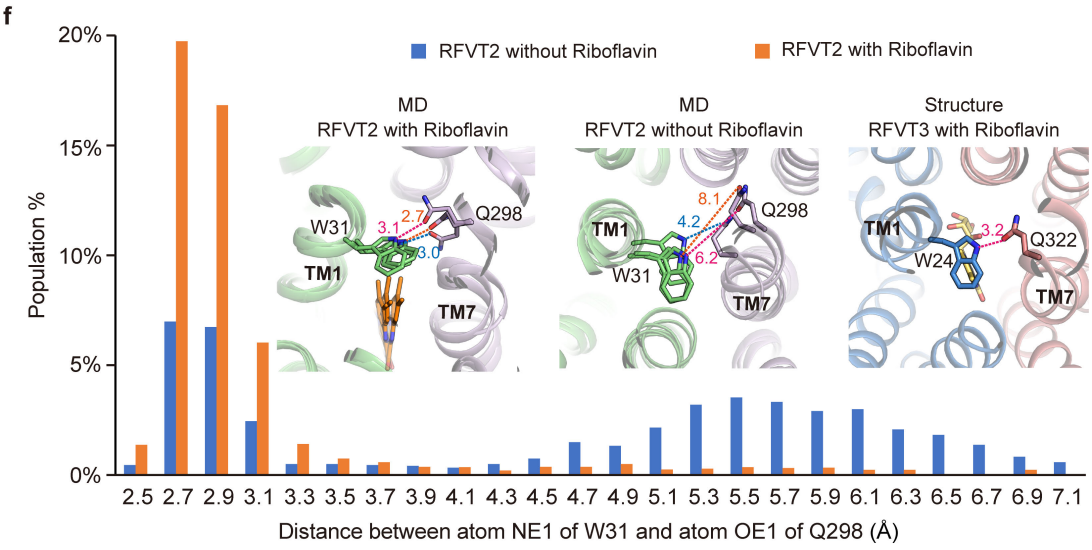


Supplementary Fig. 7. MD simulation analysis of riboflavin binding in RFVTs.

a,b, The calculated binding free energy of riboflavin to RFVT2 (blue) and RFVT3 (red) obtained by applying MM/GBSA to three independent 400 ns MD simulation trajectories. **c,e,** Per residue contribution to the MM/GBSA binding free energy of riboflavin in RFVT2 (**c**) and RFVT3 (**e**). Data represent as the mean ± standard deviation (SD) of the three independent production runs. **d,f,** The key residues in the pocket of RFVT2 (**d**) and RFVT3 (**f**) were depicted as sticks and color-coded by their contributions to the MM/GBSA binding free energy, with red indicating high contribution and white indicating low contribution. **g,** Root-Mean-Square Fluctuations (RMSF) of non-hydrogen atoms of riboflavin. RMSF calculations were conducted on the concatenated trajectory of three independent 400-ns production run. To ensure equilibrium, the first 50 ns of each run were excluded before concatenation. The reference structure used for RMSF calculation was the average structure obtained by aligning protein backbones in the concatenated trajectory to its first frame. Source data are provided as a Source Data file.

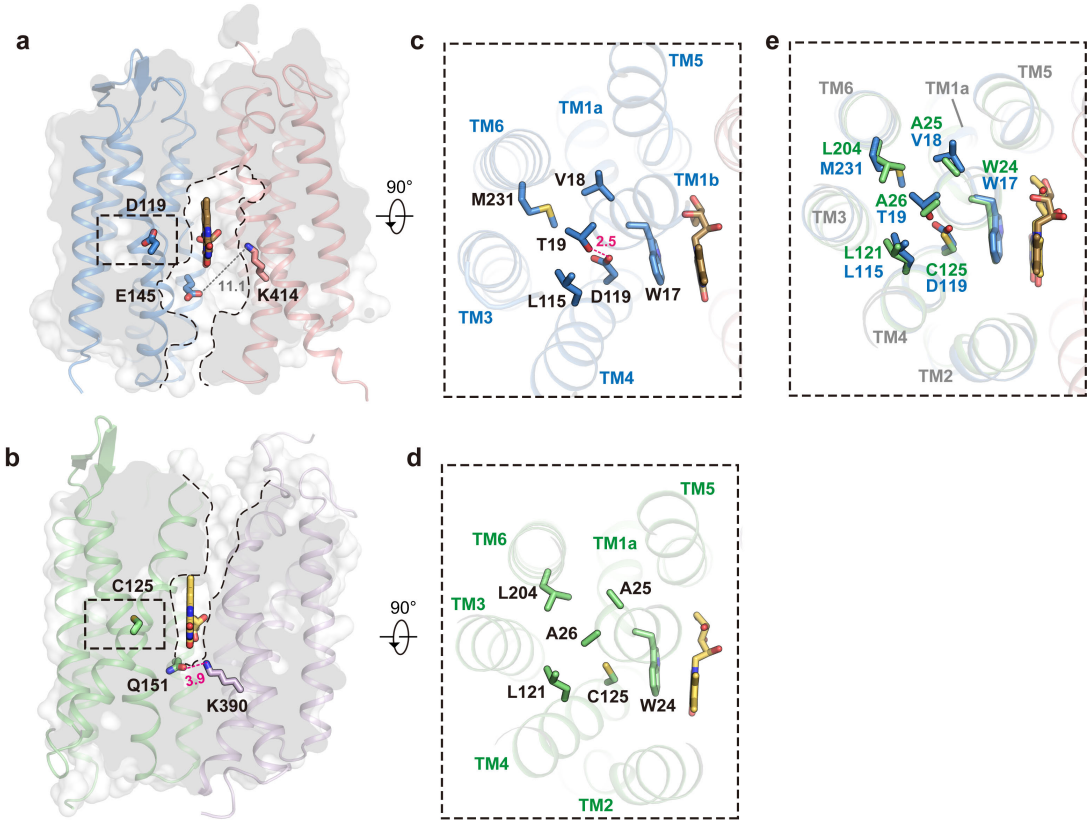

a
RFVT2 with Riboflavin
RMSD (Å)
Tims (ns)
b
RFVT3 with Riboflavin
RMSD (Å)
Tims (ns)
c
RFVT2 without Riboflavin
RMSD (Å)
Tims (ns)
d
RFVT3 without Riboflavin
RMSD (Å)
Tims (ns)
Run 1
Run 2
Run 3


e
OF RFVT2 with Riboflavin
OF RFVT2 without Riboflavin
IF RFVT3 with Riboflavin
IF RFVT3 without Riboflavin
RFVT2
RFVT3
Out
In
TM1
TM2
TM3
TM4
TM5
TM6
TM7
TM9
TM10
TM11
Riboflavin
Loop10-11
Loop4-5
10.6
20.4
Population %
Distance between the centroids of Loop4-5 and Loop10-11 (Å)


f
RFVT2 without Riboflavin
RFVT2 with Riboflavin
MD
RFVT2 with Riboflavin
MD
RFVT2 without Riboflavin
Structure
RFVT3 with Riboflavin
W31
Q298
TM1
TM7
W24
Q322
Population %
Distance between atom NE1 of W31 and atom OE1 of Q298 (Å)

Supplementary Fig. 8. MD simulation analysis of RFVTs in inward-facing and outward-occluded states.

a-d, The protein backbone RMSD plots for MD simulation systems of RFVT3 with riboflavin (**a**), RFVT2 with riboflavin (**b**), RFVT3 without riboflavin (**c**) and RFVT2 without riboflavin (**d**). The protein backbones in each trajectory were aligned with the first frame prior to calculating the RMSD. The RMSD plots of the three independent 400-ns production runs are colored in blue, gray, and orange, respectively. **e,** Histogram distribution of the distance between the centroids of Loop10-11 and Loop4-5 during MD simulations. The distance between centroids of the loop of TM10 and TM11 (Loop10-11: V420 to A428 in RFVT3; L396 to A404 in RFVT2) and the loop of TM4 and TM5 (Loop4-5: M130 to L138 in RFVT3; L136-L144 in RFVT2) were used to indicate the inward-facing and outward-occluded states of RFVTs. A smaller distance between these centroids indicated an outward-occluded state, while an increase in the distance implies a transition from outward-occluded to inward-open state. The distribution for outward-facing (OF) system (purple and black) are considered as the benchmark of outward-occluded state. Compared to the inward-facing (IF) with riboflavin (green), the IF without riboflavin (yellow) shows a distribution notably closer to that of outward-occluded state. To calculate the histogram distribution, three independent 400-ns production runs were conducted for each system, and their trajectories were concatenated by system. The calculation of the distance between the centroids of Loop10-11 and Loop4-5 involved considering all atoms within the loops of interest. Inlet: illustration of Loop10-11 and Loop4-5 in RFVT2 (left) and RFVT3 (right). RFVT2 in purple, RFVT3 in green, the loops in red. **f,** Histogram distribution of the distance between atom NE1 of W31 and atom OE1 of Q298 in RFVT2 during MD simulations. Three independent 400-ns production runs were conducted for two systems: RFVT2 with and without Riboflavin. For each system, the trajectories were concatenated and then subjected to histogram distribution analysis. Inlet: representative structures selected from the simulation trajectories of RFVT2 with riboflavin (left) and without riboflavin (middle), demonstrating the formation of hydrogen bond between W31 and Q298; the distance between the equivalent W24 and Q322 in the inward-facing RFVT3 (right). Representative structures were selected using a protein backbone-based clustering approach with a 1.5 Å cutoff using the GROMOS algorithm. The average structure of the cluster with the highest population was chosen as the representative structure for each trajectory. Source data are provided as a Source Data file.



Supplementary Fig. 9. D119 and E145 of RFVT3 are not conserved in RFVT2.

a,b, The cut-open surface shows the positions of D119 and E145 in the cavity of RFVT3 (**a**) and C125 and Q151 in the cavity of RFVT2 (**b**). Riboflavin and side chains are shown in sticks. **c,d,** D119 in RFVT3 (**c**) and C125 in RFVT2 (**d**) is enclosed in an extended pocket in the N-domain. **e,** The superposition of C125 in RFVT2 and D119 in RFVT3. The N-domain was used as a reference for structural alignment. Side chains and riboflavin are depicted in sticks. Red dashed lines represent hydrogen bond.

Supplementary Table 1. Cryo-EM data collection, refinement and validation statistics.

	RFVT2-riboflavin (EMD-38622) (PDB: 8XSM)	RFVT3-riboflavin (EMD-38623) (PDB: 8XSN)
Data collection and processing		
Magnification	105,000 ×	105,000 ×
Voltage (kV)	300	300
Electron exposure (e^{-}/Å^2)	60	60
Defocus range (μm)	–1.0 ~ –2.0	–1.0 ~ –2.0
Pixel size (Å)	0.85	0.85
Symmetry imposed	C1	C1
Initial particle images (no.)	2,720,825	4,273,317
Final particle images (no.)	143,636	99,741
Map resolution (Å)	3.01	3.29
FSC threshold	0.143	0.143
Map resolution range (Å)	2.5 ~ 4.5	2.8 ~ 4.8
Refinement		
Initial model used	AF2 model	AF2 model
Model resolution (Å)	3.13	3.40
FSC threshold	0.5	0.5
Map sharpening *B* factor (Å^2)	–121.6	–149.5
Model composition		
Non-hydrogen atoms	2,788	2,703
Protein residues	375	357
Ligands	2	1
B factors (Å^2)		
Protein	54.97	60.74
Ligand	58.06	41.46
R.m.s. deviations		
Bond lengths (Å)	0.004	0.006
Bond angles (°)	0.686	0.809
Validation		
MolProbity score	1.83	1.66
Clashscore	3.89	6.25
Poor rotamers (%)	0.70	0.68
Ramachandran plot		
Favored (%)	96.21	95.44
Allowed (%)	3.79	4.56
Disallowed (%)	0.00	0.00

Supplementary Table 2. Primer sequences.

Primers	Forward sequences	Reverse sequences
RFVT2^{WT}	CAATTCAAAGGCCTACGTCGACATGGCA GCACCCACGCCCGCCCGTCCG	GGCTCTAGATCATGCGGCCGCTCAGGAG TCACAGGGGTCTGCACAGT
RFVT2^{W24A}	GCTCCGCGGCTGCGGTCAATGGGATCT G	ATTGACCGCAGCCGCGGAGCCCATGCC GAAGAG
RFVT2^{W24F}	GCTCCTTCGCTGCGGTCAATGGGATCTG GG	ATTGACCGCAGCGAAGGAGCCCATGCCG AAGAG
RFVT2^{N28A}	GGTCGCTGGGATCTGGGTGGAGCTACC TG	CACCCAGATCCCAGCGACCGCAGCCCA GGAGC
RFVT2^{W31A}	GGGATCGCGGTGGAGCTACCTGTGGTG GTC	GTAGCTCCACCGCGATCCCATTGACCGC AGCC
RFVT2^{W31F}	GGGATCTTCGTGGAGCTACCTGTGGTG GTC	GTAGCTCCACGAAGATCCCATTGACCGC AGCC
RFVT2^{W45A}	CCAGAGGGTGCGAGCCTCCCCTCTTAC GTC	GGGGAGGCTCGCACCCTCTGGAAGCTCT TTGA
RFVT2^{E77A}	GGAAAGGACGCGCAGGTCCCCATCCGG GTGGTG	GGACCTGCGCGTCCTTTCCTGGGGCCAG CC
RFVT2^{E77R}	GGAAAGGACCGGCAGGTCCCCATCCGG GTGGTG	GGGACCTGCCGGTCCTTTCCTGGGGCCA GCCT
RFVT2^{N129A}	TGCCTCGGCCGTCACTTTCCTGCCCTTC TTG	GAAAGTGACGGCCGAGGCACAGCATGC CAGTG
RFVT2^{Q151A}	CCTGGGTGCAGGCCTGAGTGCCCTGCT GCCC	ACTCAGGCCTGCACCCAGGAAGAATGAC CGTA
RFVT2^{N291A}	GCTGACCGCCGGCGTGCTGCCTGCCGT GCA	CAGCACGCCGGCGGTCAGCGCGTTGGT GGCGGCCAA
RFVT2$^{F300A/L303A}$	GCAGAGCGCCTCCTGCGCCCCCTACGG GCGTCTGGCCT	CCCGTAGGGGGCGCAGGAGGCGCTCTG CACGGCAGGCAGCA
RFVT2^{K390A}	CTACGTGGCGGTGGCAGCCAGCTCCCT GCTG	TGGCTGCCACCGCCACGTAGGAGAACAC GCCAA
RFVT2^{K390R}	CTACGTGGCGGTGGCAGCCAGCTCCCT GCTGC	TGGCTGCCACCGCCACGTAGGAGAACAC GCCAAG
RFVT2^{K390Q}	CTACGTGCAGGTGGCAGCCAGCTCCCT GCTG	TGGCTGCCACCTGCACGTAGGAGAACAC GCCAA
RFVT2^{R402A}	GCGGGGGCGCCCCGGCATTGCTGGCAG CCG	AATGCCGGGGCGCCCCCGCCATGCAGC AGGGA
RFVT2^{R402E}	GCGGGGGCGAACCGGCATTGCTGGCAG CCG	AATGCCGGTTCGCCCCCGCCATGCAGCA GGGA
RFVT2^{Q413A}	TGGCCATCGCGGTGGGCTCTCTGCTCG GCGCT	GAGCCCACCGCGATGGCCACGCCGGCT GCCAG
RFVT2^{Q413N}	TGGCCATCAACGTGGGCTCTCTGCTCGG CGC T	GAGCCCACGTTGATGGCCACGCCGGCT GCCAG
RFVT3^{WT}	AATTCAAAGGCCTACGTCGACATGGCCT TCCTGATGCACCTG	GCTCTAGATCATGCGGCCGCTCAGGCTG GACAGTGCAGATTGCAG
RFVT3^{D119C}	CCCTGGTGTGTTGCACCTCTTCAGTGAC CTTCCT	AAGAGGTGCAACACACCAGGGCCAGGAA GAAGG
RFVT3^{E145Q}	TTGTGGGTCAAGGACTCAGCGGCCTCTT GC	GCTGAGTCCTTGACCCACAAAGAAGGTG GTGAG